

Definición de la infraestructura para procesos masivos de argumentación mediante aplicación de revisión de creencias y argumentación sobre ontologías Datalog+/-

Cristhian A. D. Deagustini María Vanina Martínez Marcelo A. Falappa
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingenierías de la Computación
Universidad Nacional del Sur
Alem 1253 - Bahía Blanca - Buenos Aires - Argentina
(0291) 459-5135

cadd@cs.uns.edu.ar, vanina.martinez@cs.ox.ac.uk, mfalappa@cs.uns.edu.ar, grs@cs.uns.edu.ar

Resumen

En los últimos tiempos, la colaboración y el intercambio de información se han vuelto aspectos cruciales de muchos sistemas. En estos entornos es de vital importancia definir métodos automáticos para resolver conflictos entre el conocimiento compartido por distintos sistemas. Este conocimiento es frecuentemente expresado a través de ontologías que pueden ser compartidas por los sistemas que utilizan el mismo.

En la presente investigación se busca la definición de métodos automáticos de resolución de conflictos en ontologías Datalog+/- . En base a lo logrado en este aspecto se buscará la adaptación del framework desarrollado para su aplicación tanto en la creación de federaciones de Bases de Datos (Data Federation) como en el intercambio de datos (Data Exchange). En estos campos de aplicación estos métodos podrán contribuir brindando la posibilidad de obtener de forma automática un esquema universal que respete tanto como sea posible a los originales manteniendo la coherencia del mismo con respecto a las restricciones de integridad impuestas a los datos, y definiendo que datos pueden ser mantenidos en la federación resolviendo incoherencias en el proceso.

Adicionalmente, se analizarán posibles extensiones a Datalog+/- basadas en formalismos de Argumentación Rebatible, teniendo en cuenta aspectos como la definición de relaciones de inferencia para estas ontologías aumentadas que tengan en cuenta los aspectos no-monótonos de la Argumentación Rebatible, o el impacto de tales relaciones en las conclusiones finales obtenidas y la complejidad de la obtención de las mismas.

Palabras Clave: Integración de Bases de Conocimiento, Revisión de Creencias, Representación de Conocimiento, Razonamiento, Argumentación Rebatible.

1. Contexto

Esta línea de investigación se lleva a cabo en el marco de los siguientes proyectos de investigación:

- **“Representación de conocimiento y razonamiento argumentativo: Herramientas inteligentes para la web y las bases de conocimiento”**. Director: Guillermo R. Simari. En evaluación, 01/01/15 – 31/12/2018. Unidad coordinadora: Universidad Nacional del Sur.
- **“Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente”**. Director: Marcelo A. Falappa. PIP 112-20110101000. Unidad coordinadora: Consejo Nacional de Investigaciones Científicas y Técnicas.

2. Introducción

En los últimos tiempos la integración e interacción entre diferentes sistemas se ha vuelto muy común, especialmente en entornos colaborativos introducidos desde el arribo de la Web Semántica [BLHL01], *e. g.*, e-commerce.

Sin embargo, la colaboración entre sistemas no siempre puede realizarse de manera directa. Hay ocasiones donde problemas de inconsistencia (e incoherencia) aparecen cuando tomamos el conocimiento provisto por diferentes entidades como uno solo. En formalismos de representación de conocimiento como las ontologías estas inconsistencias suelen aparecer como violaciones a las restricciones de integridad impuestas a los datos en las fuentes originales del conocimiento, que son violadas en el conocimiento integrado.

La resolución de incoherencias e inconsistencias en conocimiento es admitido como un problema importante

que debe ser atacado [GCS10, BQL07], especialmente en procesos de integración de conocimiento proveniente de fuentes diversas [BHP09, AK05]. En particular; el problema de incoherencia es conocido en formalismos de representación de conocimiento, especialmente en la comunidad de Lógicas Descriptivas (Description Logics - DL), donde ha sido analizado desde diferentes enfoques a través de los años [FHP⁺06, BB97]. En base a tales esfuerzos se han desarrollado adaptaciones del concepto para Datalog+/- [DMFS, DMFS15], que han sido de gran importancia para el desarrollo de métodos de integración que atiendan conflictos en el conocimiento ontológico además de aquél expresado en bases de datos.

La obtención de conocimiento coherente y consistente basado en distintas fuentes es un problema que aparece en muchos campos de la IA, *e. g.*, sistemas multi-agentes, y fue atacado desde diferentes enfoques. Un área que ha hecho muchos avances en la integración de fuentes inconsistentes de información es la Revisión de Creencias a través de la definición de procesos de Integración de Creencias (Belief Merging [BHA⁺01]). A través de los años, diferentes enfoques para la integración se han desarrollado, *e. g.*, [BKM91, Cho98, LS98, FKIRS12]. En un proceso de integración se toman varias Bases de Conocimiento (Knowledge Bases - KBs) que pueden ser consistentes en sí mismas pero generan conflicto cuando son consideradas juntas, y se obtiene una nueva KB que está libre de conflictos y refleja la información de las bases originales tanto como sea posible. Esta nueva base de conocimiento puede luego ser explotada mediante procesos de inferencia clásicos, los cuales no podrían ser aplicados directamente sobre la unión de las ontologías originales ya que los mismos no toleran inconsistencias.

Alternativamente, es posible tomar un camino diferente, y convivir con la inconsistencia, tratándola de manera local a las consultas que se intente resolver, en un proceso *ad-hoc*. Esto es, en lugar de resolver cada conflicto y limpiar la base de conocimiento para luego aplicar técnicas clásicas de inferencia, es posible *modificar cómo las inferencias son realizadas*, de forma que las respuestas dadas sean consistentes. Un área muy importante que maneja tal enfoque es la de Argumentación [RSvB09, GS04, GCS10].

Los métodos presentados suelen diferir en gran manera en la forma en que encaran los conflictos en el conocimiento; sin embargo, sin importar si se enfocan en eliminar los conflictos en la base de conocimiento o modificar cómo las respuestas son obtenidas para tolerar los conflictos sin resolverlos todos coinciden en que la estrategia debe ser de forma tal que la calidad de la información sea comprometida en la menor medida posible.

En esta línea de investigación encaramos la resolución de conflictos en ontologías a través de ambos enfoques. Esto es, utilizaremos tanto la consolidación e integración de ontologías desarrolladas en Datalog+/- [CGL12] como la definición de extensiones al lenguaje que permitan el uso de información rebatible [Pol87, SL92] a través de argumentación [RSvB09, GS04, GCS10], donde las técnicas a desarrollar se guiarán principalmente por este

concepto de preservación de la calidad de información, así como también por la eficiencia de los procesos utilizados; desarrollando técnicas que se adecuen a su uso en formalismos de representación de conocimientos aptos para el uso en entornos colaborativos como la Web Semántica. En los últimos años, las bases de conocimiento en forma de ontologías se han vuelto muy populares en estos entornos, ya que proveen formas de representar tanto los datos disponibles en sí como las restricciones impuestos a éstos. Además, el poder expresivo de las ontologías permite realizar tareas importantes en la integración de fuentes de conocimientos [Len02], y juega un rol preponderante en la Web Semántica[BLHL01].

Por lo tanto, es importante definir métodos que permitan el uso de ontologías resolviendo todos los problemas de incoherencia/inconsistencia que puedan aparecer. Esto podría ser el primer paso para la integración de otras fuentes de conocimiento que pueden ser expresadas mediante estas ontologías (*e. g.*, bases de datos relacionales). Es por esto que en esta línea de investigación nos enfocamos en la definición de la infraestructura necesaria para la realización de procesos masivos de argumentación, donde tales procesos se soportarán sobre ontologías desarrolladas en un lenguaje particular denominado Datalog+/- [CGL12]. La familia de lenguajes de ontologías Datalog permite un estilo modular de representación de conocimiento mediante el uso de reglas de forma similar a la usada en Programación Lógica, y su decidibilidad le permite manejar los volúmenes masivos de datos que podemos encontrar en aplicaciones hoy en día, haciéndola útil en diferentes campos como la consulta de ontologías, extracción de datos en web o intercambio de datos [LMS12]. En particular, la representación de conocimiento en ontologías Datalog+/- se lleva a cabo mediante el uso de (a) una Base de Datos: un conjunto de átomos que representan hechos acerca del mundo, *e. g.*, alumno(pedro) (b) Tuple-generating Dependencies - TGDs: reglas que nos permiten obtener nuevos átomos mediante la activación de las mismas como ser: $\text{alumno}(X) \rightarrow \text{persona}(X)$, (c) Equality-generating Dependencies: reglas que restringen la generación de átomos, por ejemplo: $\text{doctor}(D, P) \wedge \text{doctor}(D', P) \rightarrow D = D'$; y (d) Negative Constraints - NCs: reglas que expresan relaciones que no pueden existir entre átomos, *e. g.*: $\text{alto}(X) \wedge \text{bajo}(X) \rightarrow \perp$. Los métodos a desarrollar se enfocarán, por lo tanto, en controlar la relación entre los átomos y las TGDs presentes en la ontología, de forma que se respeten las EGDs y NCs en la KB.

3. Líneas de Investigación y Desarrollo

Esta línea de investigación se enfoca en la definición de procesos de resolución de conflictos en ontologías Datalog+/- a través del uso de formalismos enfocados en la resolución de incoherencias e inconsistencias provenientes de las áreas de Revisión de Creencias y Argumentación. Para ello distintos ejes deben ser investigados,

que van desde la definición de incoherencias e inconsistencia en el entorno de ontologías Datalog+/- hasta las posibles aplicaciones que un método automático de integración de estas ontologías podría tener.

3.1. Definición de métodos de identificación de Incoherencias e Inconsistencias en Datalog+/-

Datalog+/- se ha vuelto un lenguaje muy popular en los últimos años, y numerosos estudios se han realizado acerca de sus propiedades de decibilidad y la complejidad asociada a la respuesta de consultas en estas ontologías. Sin embargo, no ha habido mucho estudio acerca de los aspectos de Representación de Conocimiento en Datalog+/- . Muy poco trabajo se ha hecho acerca de inconsistencias en ontologías Datalog+/- . Peor aún es la situación respecto del concepto de incoherencia (*i. e.*, la imposibilidad de satisfacer cierto conjunto de TGDs sin violar una NC).

Uno de los ejes de esta línea de investigación es la definición formal del concepto de incoherencia en Datalog+/- , tomando como partida esfuerzos similares que han sido realizados para otros formalismos de representación de conocimiento, principalmente Description Logics. Adicionalmente, se procederá a identificar las propiedades que llevan a que un conjunto de TGDs sea incoherente, y las que hacen que una ontología Datalog+/- se vuelva inconsistente. De esta manera se podrán identificar tales casos, lo que será el primer paso para la posterior resolución de tales problemas.

3.2. Resolución de Incoherencias e Inconsistencias en Datalog+/-

Una vez que se tiene definidos e identificados los conjuntos incoherentes de TGDs y aquellos conjuntos de átomos que provocan inconsistencias en la unión de varias ontologías Datalog+/- , se debe proceder a la resolución de estos conflictos. En esta línea de investigación esto se hará mediante la aplicación de técnicas derivadas de la Revisión de Creencias denominadas Kernel Contraction.

Este tipo de técnicas resuelve conflictos de incoherencia/inconsistencia tomando los conjuntos conflictivos mínimos y eligiendo de alguna forma que elemento remover de los mismos para solucionar el problema. En el caso de integración de ontologías Datalog+/- esto es la remoción de ciertos átomos y ciertas TGDs de la unión de todas las ontologías que se está integrando. Adicionalmente, se puede pensar en la definición de métodos de debilitamiento de reglas, en lugar de la remoción de las mismas. Esto no es una tarea trivial, ya que hay muchos aspectos a definir, por ejemplo como elegir el mejor candidato entre los átomos o TGDs que pueden eliminarse, lo que a su vez lleva a definir formas (automáticas) de obtener órdenes entre los candidatos. Para esto se procederá a definir operadores de integración de ontologías Datalog+/- , así como se darán las propiedades esperadas de tales operadores y se definirán métodos de obtención

de operadores de tales características mediante Teoremas de Representación.

3.3. Posibles extensiones a Datalog+/- mediante Argumentación Rebatible

Otro aspecto a considerar dentro de la línea de investigación es como se beneficiaría Datalog+/- de otros formalismos de representación de conocimientos con propiedades diferentes a aquellas presentes en Datalog+/- . Particularmente, proponemos analizar posibles extensiones a las ontologías Datalog+/- basadas en el uso de formalismos de Argumentación Rebatible como Programación Lógica Rebatible (Defeasible Logic Programming - DeLP) [GS04].

Para tales extensiones se deberán analizar diferentes aspectos. Por ejemplo, la relación de inferencia en estas ontologías aumentadas deberá tener en cuenta los aspectos no-monótonos de la Argumentación Rebatible, llevando a que se modifique la forma en que una consulta es respondida respecto del proceso actual en Datalog+/- , ya que se deberá tener en cuenta el análisis dialéctico llevado a cabo por DeLP antes de responder la misma.

Estas modificaciones en la forma en que información es inferida traerá aparejado un impacto en las inferencias finales de las ontologías en aquellos casos en donde las ontologías no son coherentes o consistentes, proveyendo otra forma de integrar las mismas (*i. e.*, simplemente considerarlas juntas sin importar problemas de incoherencia e inconsistencia, y dejar que el proceso argumentativo los resuelva).

3.4. Integración de otras fuentes de datos a través del Merging de Ontologías Datalog+/-

Finalmente, se analizará nuestro framework de integración de ontologías Datalog+/- como medio de integración de otras fuentes de datos. Principalmente nos enfocaremos en la integración automática de Bases de Datos Relacionales.

Para esto, primeramente definiremos métodos para expresar bases de datos relacionales a través de ontologías Datalog+/- , tanto los datos en sí como aspectos relacionados al esquema de las mismas, *e. g.*, las dependencias funcionales.

Una vez logrado esto se podría utilizar los métodos de integración de ontologías Datalog+/- para obtener una federación de las bases de datos expresadas, ya que la ontología final resultante de la integración brindaría un esquema integrador de las mismas así como los datos que serían parte de la federación, manteniendo a su vez la coherencia de las restricciones de integridad respecto del esquema unificado y la consistencia de los datos almacenados respecto de las dependencias funcionales.

4. Resultados y Objetivos

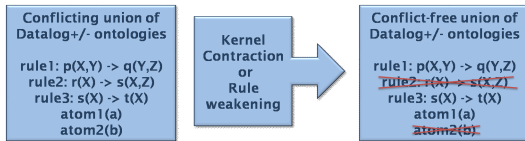
El objetivo general de este trabajo de investigación es el diseño y construcción de la infraestructura necesaria para la realización de procesos de argumentación en una escala masiva valiéndose de datos almacenados en Bases de Datos heterogéneas integradas en Federaciones.

En el presente trabajo se propone la representación de tales bases de datos a través de la utilización de ontologías Datalog+/-, las cuales poseen una expresividad y una tratabilidad adecuadas para tal fin. De esta forma, todos los desarrollos que se logren en Datalog+/- serían transferibles a bases de datos (y otras tecnologías estrictamente menos expresivas que Datalog+/- como ciertas Description Logics, *e. g.*, DL-Lite o \mathcal{EL}).

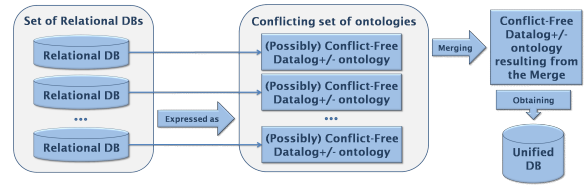
Hay dos aspectos principales a tener en cuenta para la consecución del objetivo planteado de realización de procesos masivos de argumentación. Por un lado tenemos la definición de métodos automáticos que permitan la creación de vistas unificadas de varias bases de datos como forma de construir las federaciones de bases de datos, de forma que los procesos de argumentación accedan a las mismas para dar soporte a los argumentos. Por el otro, es posible modificar la forma en que la información es inferida en Datalog+/-, de forma que la misma sea tolerante a inconsistencias e incoherencias. Debido a esta dualidad respecto a la consecución del objetivo general se plantean objetivos específicos que abarcan dos temáticas diferentes aunque relacionadas.

Respecto de la definición de métodos automáticos como forma de construir federaciones:

- Definición de métodos de consolidación de ontologías Datalog+/- que permitan el manejo de inconsistencia e incoherencia aprovechando tanto enfoques ya clásicos en la literatura de revisión de creencias como Kernel Contraction [Han94, Han99] como así también refinamientos desarrollados en la presente línea de investigación (Cluster Contraction [DMFS14b, DMFS14a]).



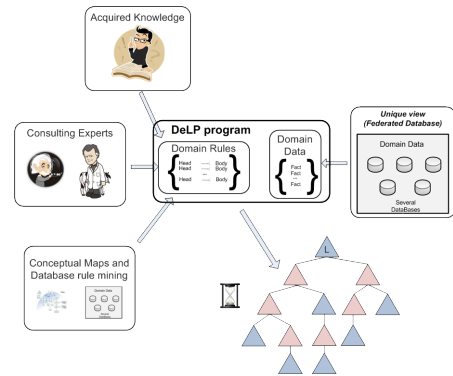
- En base a los métodos desarrollados procederemos a definir métodos que permitan la integración de ontologías Datalog+/-, mediante la utilización de la consolidación de la unión de ontologías Datalog+/- . Una vez logrado esto se podría utilizar los métodos de integración para obtener una federación de las bases de datos expresadas, ya que la ontología final resultante de la integración brindaría un esquema integrador de las mismas así como los datos que serían parte de la federación, manteniendo a su vez la coherencia de las restricciones de integridad respecto del esquema unificado y la consistencia de los datos almacenados respecto de las dependencias funcionales, como se muestra a continuación.



A su vez, respecto de la modificación de las relaciones de inferencia utilizadas en Datalog+/- para hacerlas tolerantes a inconsistencia e incoherencia tenemos:

- Desarrollo y análisis de posibles extensiones a la familia de lenguajes Datalog+/- basadas en el uso de formalismos de Argumentación Rebatible como Programación Lógica Rebatible (Defeasible Logic Programming - DeLP) [GS04], avanzando sobre los resultados introducidos por esta línea de investigación [MDFS14, DMFS15]. Estos avances se han realizado sobre conjuntos restringidos de ontologías Datalog+/- . Como continuación de tales trabajos se procederá a generalizar tales desarrollos para la totalidad del conjunto de ontologías Datalog+/- .
- Análisis del incremento que tales desarrollos provoquen en la cantidad de información que puede ser inferida de las ontologías Datalog+/- (un primer avance al respecto puede verse en [DMFS15]) como así también del trade-off que el uso de argumentación en Datalog+/- induce respecto del aumento de inferencias y la complejidad y eficiencia en la resolución de consultas.

Finalmente, se pretende analizar el uso de los métodos desarrollados en distintas aplicaciones como sistemas de soporte de decisiones (Decision Support Systems - DSS) y sistemas de recomendación (Recommender Systems - RS). Un esquema general que tales aplicaciones pueden seguir es mostrado a continuación.



5. Formación de Recursos Humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de posgrado en el Doctorado en Ciencias de la Computación del Departamento de Ciencias e Ingenierías de la Computación de la Universidad Nacional del Sur.

Referencias

- [AK05] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *ECSQA-RU 2005*, pages 527–538, 2005.
- [BB97] D. Beneventano and S. Bergamaschi. Incoherence and subsumption for recursive views and queries in object-oriented data models. *Data Knowl. Eng.*, 21(3):217–252, 1997.
- [BHA⁺01] I. Bloch, A. Hunter, A. Appriou, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. M. Cooke, F. Cuppens, D. Dubois, H. Fargier, M. Grabisch, R. Kruse, J. Lang, S. Moral, H. Prade, A. Saffiotti, P. Smets, and C. Sossai. Fusion: General concepts and characteristics. *Int. J. Intell. Syst.*, 16(10):1107–1134, 2001.
- [BHP09] E. Black, A. Hunter, and J. Z. Pan. An argument-based approach to using multiple ontologies. In *SUM*, pages 68–79, 2009.
- [BKM91] C. Baral, S. Kraus, and J. Minker. Combining multiple knowledge bases. *IEEE Trans. Knowl. Data Eng.*, 3(2):208–220, 1991.
- [BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):3443, 2001.
- [BQL07] D. A. Bell, G. Qi, and W. Liu. Approaches to inconsistency handling in description-logic based ontologies. In *OTM Workshops (2)*, pages 1303–1311, 2007.
- [CGL12] A. Cali, G. Gottlob, and T. Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.
- [Cho98] Laurence Cholvy. Reasoning about merged information. In *Belief Change*, volume 3, pages 233–263. Springer Netherlands, 1998.
- [DMFS] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Datalog+/- ontology consolidation. *J. Artif. Intell. Res. (JAIR)*. Under Review.
- [DMFS14a] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Improving inconsistency resolution by considering global conflicts. In *SUM*, pages 120–133, 2014.
- [DMFS14b] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. Inconsistency resolution and global conflicts. In *ECAI*, pages 991–992, 2014.
- [DMFS15] C. A. D. Deagustini, M. V. Martinez, M. A. Falappa, and G. R. Simari. On the influence of incoherence in inconsistency-tolerant semantics for datalog+/- . In *ONTOLP ’15, IJCAI-2015, Buenos Aires, Argentina, Julio, 2015, Proceedings*, 2015.
- [FHP⁺06] G. Flouris, Z. Huang, J. Z. Pan, D. Plexousakis, and H. Wache. Inconsistencies, negations and changes in ontologies. In *AAAI*, pages 1295–1300. AAAI Press, 2006.
- [FKIRS12] M. A. Falappa, G. Kern-Isberner, M. Reis, and G. R. Simari. Prioritized and non-prioritized multiple change on belief bases. *J. of Philos. Logic*, 41(1):77–113, 2012.
- [GCS10] S. A. Gómez, C. I. Chesñevar, and G. R. Simari. Reasoning with inconsistent ontologies through argumentation. *Appl. Artif. Intell.*, 24(1&2):102–148, 2010.
- [GS04] A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *TPLP*, 4(1-2):95–138, 2004.
- [Han94] S. O. Hansson. Kernel contraction. *J. Symb. Log.*, 59(3):845–859, 1994.
- [Han99] S. O. Hansson. *A Textbook of Belief Dynamics*. Kluwer Academic Publishers, 1999.
- [Len02] M. Lenzerini. Data integration: A theoretical perspective. In *PODS*, pages 233–246, 2002.
- [LMS12] T. Lukasiewicz, M. V. Martinez, and G. I. Simari. Inconsistency handling in datalog+/- ontologies. In *Proc. of ECAI*, pages 558–563, 2012.
- [LS98] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Trans. Knowl. Data Eng.*, 10(1):76–90, 1998.
- [MDFS14] M. V. Martinez, C. A. D. Deagustini, M. A. Falappa, and G. R. Simari. Inconsistency-tolerant reasoning in datalog+/- ontologies via an argumentative semantics. In *IBERAMIA ’14*, pages 15–27, 2014.
- [Pol87] John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.
- [RSvB09] I. Rahwan, G. R. Simari, and J. van Benthem. *Argumentation in artificial intelligence*, volume 47. Springer, 2009.
- [SL92] G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artif. Intell.*, 53(2-3):125–157, 1992.